

動画から文字情報を抽出する方法及び装置

発明の技術的背景

本発明は、動画に含まれる文字情報を抽出する方法及び装置に関する。

通常用いられている複写機や読み取スキャナが備える読み取り手段は、原稿を正確に再現するため、原稿を固定し、原稿画面をキャリッジミラーで頁面に平行に走査しながら読み込む。あるいは、スルーリード式で原稿紙を1枚ずつ紙送りして固定ミラーで原稿画像を反射し、レンズで集光してライン型CCD撮像素子で線画情報を順次メモリ内に格納し、線画情報をメモリ内でつなぎ合わせて頁画像を再構築し、この画像を電子データ化したり印刷出力を行う。

しかし、これらの装置では1枚ずつの原稿しか読み取りができず、書籍のような頁数が多く綴じられた原稿を読み取ることができなかった。

特開平9-200451号公報は、書籍のような原稿を読み取ることが可能であって、頁間の画像濃度を比較して頁の切り替わりを検知する装置を提案している。

また、特開平2000-201358号公報は、動画を構成する各々の静止画をつなぎ合わせて、1つのパノラマ画像に統合する映像記録装置を提案している。

しかし、書籍に含まれる文字、あるいは撮影した動画映像に含まれる文字を効率良く抽出し、後に加工が可能なように文書データを生成する技術は存在しなかった。

発明の概要

そこで本発明は、書籍や動画映像に含まれる画像情報のうち、後に利用性の高い文字情報に注目し、動画映像から文字領域を識別し、この文字領域における画像情報を文字情報に変換して加工性の良い文書データを出力することが可能な装置及び方法を提供することを目的とする。

本発明によれば、文字を含む撮影対象物を撮影して動画情報を生成するステップと、前記動画情報に含まれる静止画を取り出すステップと、前記静止画に含まれる文字領域を識別するステップと、識別した前記文字領域の画像情報を文字情報に変換するステップとを備える、動画から文字情報を抽出する方法が提供される。

ここで、前記撮影対象物を撮影して前記動画情報を生成するステップは、前記撮影対象物が原稿台に載置されているか否かを判定するステップと、前記撮

影対象物が載置されていない場合に、操作者に前記撮影対象物を載置することを促す表示を行うステップと、前記原稿台に載置された前記撮影対象物を撮影して前記動画情報を生成するステップとを備えてもよい。

また、前記動画情報に含まれる前記静止画を取り出すステップは、前記動画情報に含まれる画像の動き率が所定値以下である静止画を取り出すステップと、取り出した前記静止画をそれぞれメモリに格納するステップとを備えてもよい。

前記メモリは、外部に着脱可能な、コンピュータにより読み取り可能な記録媒体であってもよい。

前記静止画に含まれる前記文字領域を識別するステップは、前記文字領域に対して文字の認識が可能か否かを判断し、可能でない場合、文字の認識が可能になるまで、撮影中である場合は撮影装置のズーム率を高めていき、撮影済みである場合は撮影後の前記静止画のズーム率を高めていくステップと、前記ズーム率を最大にしても文字の認識が可能にならない場合、前記静止画に含まれる前記文字領域及び非文字領域とを統合した画像情報を生成するステップとを有し、識別した前記文字領域における前記画像情報を前記文字情報に変換するステップは、前記文字領域に対して文字の認識が可能である場合、前記文字領域にOCR処理を施し、前記文字領域における前記画像情報を前記文字情報に変換するステップを有することもできる。

前記撮影装置のズーム率を高めていくステップは、前記ズーム率を高めた後、横方向及び／又は縦方向の端部を検出するまで画像を移動していき、前記文字領域が存在するか否かを判断し、前記文字領域が存在する場合、識別した前記文字領域における前記画像情報を前記文字情報に変換するステップへ移行するように構成してもよい。

また本発明における、ネットワークを利用した動画から文字情報を抽出する方法は、文字を含む撮影対象物を撮影して動画情報を生成するステップと、前記ユーザが前記動画情報を通信ネットワークを介してサービス提供者に送信するステップと、前記サービス提供者が、送信された前記動画情報に含まれる静止画を取り出すステップと、前記静止画に含まれる文字領域を識別するステップと、識別した前記文字領域の画像情報を文字情報に変換するステップと、変換された前記文字情報を前記通信ネットワークを介して前記ユーザに送信し、あるいは前記文字情報が格納された記録媒体を前記ユーザに送付するステップとを備えている。

本発明の動画から文字情報を抽出する装置は、文字を含む撮影対象物を撮影して動画情報を生成する撮影装置と、前記動画情報に含まれる静止画を取り出す静止画抽出部と、前記静止画に含まれる文字領域を識別する文字領域識別部

と、識別した前記文字領域の画像情報を文字情報に変換する文字情報変換部とを備えている。

ここで、前記静止画抽出部は、前記動画情報に含まれる画像の動き率が所定値以下である静止画を取り出す画像動き率判定部と、取り出した前記静止画をそれぞれ格納するメモリとを備えてもよい。

前記メモリは、外部に着脱可能な、コンピュータにより読み取り可能な記録媒体であってもよい。

本発明における、ネットワークを利用した動画から文字情報を抽出する装置は、ユーザ側において、文字を含む撮影対象物を撮影して動画情報を生成する撮影装置と、前記ユーザが前記動画情報を通信ネットワークを介してサービス提供者に送信する送信装置と、前記サービス提供者側において、送信された前記動画情報に含まれる静止画を取り出す静止画抽出部と、前記静止画に含まれる文字領域を識別する文字領域識別部と、識別した前記文字領域の画像情報を文字情報に変換する文字情報変換部と、変換された前記文字情報を前記通信ネットワークを介して前記ユーザに送信する送信装置とを備えている。

前記静止画抽出部は、前記動画情報に含まれる画像の動き率が所定値以下である静止画を取り出す画像動き率判定部と、取り出した前記静止画をそれぞれ格納するメモリとを備えてもよい。

ここで前記メモリは、外部に着脱可能な、コンピュータにより読み取り可能な記録媒体であってもよい。

図面の簡単な説明

添付図面において、

図1は、本発明の実施例による動画から文字情報を抽出する装置の構成を示すブロック図。

図2は、同装置を用いて文字情報の抽出が可能な、文字情報及び画像情報を有する原稿を示す説明図。

図3A、3Bは、本発明の実施例による動画から文字情報を抽出する方法における処理の手順を示したフローチャート。

図4Aは、同方法において、撮影して得られた映像信号に画像処理を施してメモリに格納する手順を示したフローチャート。

図4Bは、同方法において、動画から静止画を抽出する処理の手順を示したフローチャート。

図5は、同方法において、操作者に原稿設置を促すときの画面の表示を示したフローチャート。

図6は、同方法において、文字を識別するためにデジタルズームを行う手順

を示したフローチャート。

図7は、同方法において、文字領域にOCR処理を行って文字を認識し、得られた文書データと、非文字領域とを統合する処理の手順を示したフローチャート。

図8は、ユーザとサービス提供者とのネットワーク接続を示した説明図。

図9は、ユーザがサービス提供者にネットワークを介してサービスの提供を求めるときの手順を示したフローチャート。

図10は、ユーザがサービス提供者に登録を行うときの手順を示したフローチャート。

好適な実施例の説明

以下、本発明の実施例について図面を参照して説明する。

先ず、図1に本実施例による動画から文字情報を抽出する装置の構成を示す。本装置は、得られた動画映像信号に対して所定の処理を行う画像処理部10と、原稿台160上に載置された原稿130を読み取る原稿読み取り装置150に含まれるカメラ140及びカメラ140に含まれるレンズの動作を制御するカメラ／レンズ制御部100とを備えている。ここで、カメラ140はスチルカメラではなく、動画映像の撮影が可能なビデオカメラである。

画像処理部10は、入力ソース判定部20、静止画抽出部30、文字領域識別部40、OCR処理部50、文字及び画像領域統合部60を有する。カメラ／レンズ制御部100は、カメラ首振り動作制御部110、ズーム及びパン動作制御部120を有する。

入力ソース判定部20は、画像処理部10に入力された動画映像信号がいずれの入力ソースによるものであるか、即ち撮影済みの動画映像信号であるか、あるいは原稿読み取り装置150を用いてこれから撮影して得られる動画映像信号であるかを判定する。

静止画抽出部30は、動画映像信号に含まれる静止画像を抽出する。ここで、入力ソースが原稿読み取り装置150によりこれから撮影して得られる動画映像信号である場合は、カメラ首振り動作制御部110によるカメラ首振り動作の制御と連携して静止画像の抽出を行う。

また、図2に示されたように、抽出した静止画像200には、多くの場合、文字領域210、220と画像領域（非文字領域）230、240とが含まれる。文字領域識別部40は、抽出した静止画像200に含まれる文字領域210、220、画像領域230、240のうち、文字領域210、220を識別する。ここで、入力ソースが原稿読み取り装置150によりこれから撮影して得られる動画映像信号である場合は、ズーム及びパン動作制御部120における

るズーム及びパン動作の制御と連携して、文字領域の識別を行う。

O C R 处理部 5 0 は、識別された文字領域に対し、O C R (optical character reader) 处理を行い、画像情報から文字情報を取得する。

文字及び画像領域統合部 6 0 は、文字情報を取得することができなかった場合（文字情報を取得できた場合を含めてもよい）、文字領域及び画像領域を統合したデータを出力する。これは、文字情報が得られたかったにもかかわらず、後に文字情報を利用しようとする場合があり得ることに備えたリスクマネジメント処理として行うものであり、解像度や再現性が低い場合であっても何らかの出力を得るために行う。

次に、このような構成を備えた本実施例における動作について、図 3 A、図 3 B のフローチャートを用いて説明する。

ステップ S 1 0 0において、先ず画像メモリをリセットする。

ステップ S 1 0 2として、画像処理部 1 0 の入力ソース判定部 2 0 が、画像処理部 1 0 に入力された動画映像信号がいずれの入力ソースによるものであるか、即ち撮影済みの動画映像信号であるか、あるいは原稿読み取り装置 1 5 0 を用いてこれから撮影して得られる動画映像信号であるかを判定する。

入力された動画映像信号が、撮影済みの信号である場合は、ステップ S 1 0 4として、動画映像信号を入力するシーケンスへ移行する。このシーケンスを開始する際には、予め図 4 A に示された手順で、動画映像信号が得られている必要がある。

ステップ S 2 0 0としてビデオカメラ等の動画映像撮影装置を用いて、文字を含む対象物を撮影し、動画映像信号を生成する。

得られた動画映像信号を、メモリやハードディスク、テープ等、コンピュータにより読み取り可能な記録媒体に一旦格納する。

ステップ S 2 0 2として、動画映像信号にノイズ除去等の所定の画像処理を行い、ステップ S 2 0 4として何等かの記録媒体に格納する。

そして、図 4 B に示された手順で、得られた動画映像信号を入力する。ステップ S 3 0 0として、動画映像信号の有無を判定し、無い場合はステップ S 3 0 0に戻る。

動画映像信号が存在する場合は、ステップ S 3 0 2に移行し、静止画像の抽出処理を行う。具体的には、画像の動き率が所定値 R（例えば、2%）以下であるか否かを判断し、所定値 R を超える場合はステップ S 3 0 2へ戻る。画像の動き率が所定値 R 以下であるとき静止画像であると判断し、ステップ S 3 0 4として得られた静止画像の画像データに番号を付与し、画像メモリに保存する。

ここで、画像メモリはコンピュータで読み取り可能な記録媒体であり、図 1

に示されたC D (Compact Disc) - R W(Rewritable) 2 0 0 やM O (Magneto Optical) ディスク等、外部へ着脱可能な記録媒体であってもよい。

ステップS 3 0 6として、仮保存した画像データに含まれる文字を認識することが可能か否かを判断し、可能でない場合はステップS 3 0 8へ移行し、図6に示されたステップS 4 0 0を実行する。文字を認識することができる場合は、ステップS 3 1 0としてこの画像データを保存し、O C R処理部5 0によるO C R処理を行う。

ここで、O C R処理は、図7に示された手順に従って行う。ステップS 5 0 0として、O C R処理部5 0により、画像データに含まれる文字を認識する。

ステップS 5 0 2として、認識した文字に基づいて、画像データを文書形式の文書データに変換する。

ステップS 5 0 4として、得られた文書データと、文字が含まれない画像領域(非文字領域)の画像データとを統合する。

図4 BのステップS 3 1 0におけるO C R処理が終了すると、ステップS 3 0 0へ戻る。

ステップS 3 0 6で、文字認識を不可能と判断し、ステップS 3 0 8へ移行してデジタルズーム処理を実行する場合について、図6を用いて述べる。

ステップS 4 0 0として、画像データに対してデジタルズーム処理を行う。

ステップS 4 0 2において、文字識別が可能な否かを判定し、可能であると判断した場合はステップS 4 0 4へ移行する。ステップS 4 0 4において、このときのデジタルズーム率に基づいて、抽出すべき画像の解像度を設定する。さらに、ステップS 4 0 6として、画像データを画像メモリに保存し、上述のO C R処理を行う。

文字識別が不可能であると判断した場合は、ステップS 4 0 8においてズーム率が最大か否かを判断し、最大でない場合はステップS 4 0 0へ戻る。ズーム率が最大であるときは、もはや文字識別は不可能であると判断し、ステップS 4 1 0として、文字及び画像領域を統合した情報として出力する。

図3 AのステップS 1 0 2において、入力ソースが、原稿読み取り装置1 5 0を用いてこれから撮影して得られる動画映像信号であると判断した場合、先ず、ステップS 1 0 6として原稿を設置する作業に入る。

ステップS 1 0 8として、原稿読み取り装置1 5 0に対して、操作者が原稿読み取りの開始を指示する操作を行う。

ステップS 1 1 0として、カメラ1 4 0が撮影して得られた動画映像信号を入力する。

ステップS 1 1 2として、入力された動画映像信号に基づいて、原稿台1 6 0上に原稿1 3 0が存在するか否かを判断する。ここで、原稿1 3 0の有無を

判断する際には、得られた動作映像信号と、カメラ140が原稿台160のみを撮影したときの映像とを比較し、相違がない場合は原稿130が存在しないと判断し、相違がある場合は原稿130が存在すると判断する。

原稿130が存在しないと判断した場合は、ステップS114として、原稿設置を促すシーケンスへ移行する。このシーケンスは、図5におけるステップ350として示されたように、操作パネルにおいて「原稿を置いて下さい。処理を中止する場合は、ストップボタンを押してください。」という旨の表示を操作者へ向けて行う。

原稿130が存在すると判断した場合は、ステップS118へ移行し、光学的なパン操作を実行する。このカメラ140のパン操作は、ズーム及びパン動作制御部120によって制御される。

ステップS120として、取り込んだ動画映像信号を用いて、原稿130のサイズをメモリに記憶する。このサイズは、動画映像信号が構成する画面枠に対する、原稿の縦横の比率として表現される。ここで、原稿130のサイズを認識する際には、原稿130の背景に位置する原稿台160を認識することによって行う。

ステップS122として、記憶した、画面枠に対する原稿の縦比率と横比率のうち、大きい比率、即ち画面枠に対して余裕の少ない方の寸法に基づいて光学的なズーム動作を行う。

ステップS124として、静止画抽出部30が、動画映像信号に含まれる静止画像を取り込む処理を行う。例えば、画像の動き率が所定値R（例えば、2%）以下である場合に静止画であると判断し、抽出する。

原稿読み取り装置150を用いて原稿130を読み取る際には、操作者が原稿130の頁をめくる毎に所定時間静止する映像の中から、静止画像のみを自動的に抽出することになる。

ここで、静止画像の抽出処理には、特開平7-23322号公報、特開平8-9314号公報等において開示されている公知技術を用いて行うことができる。例えば、動画を構成する各フレーム間の画像情報の差分を求め、差分が所定値以下である場合に静止画であると判断する。あるいは、各フレーム毎に画像が所定時間変化（移動）しない場合に静止画像であると判断する。ここで、判断の基準となる所定時間を任意の長さに設定することで、静止画像の抽出処理に自由度を持たせることができる。

ステップS126として、抽出した静止画像に番号を付与し、画像データとして画像メモリに一次的に保存する。

図3BにおけるステップS128として、仮保存した画像データに含まれる文字を認識することができるか否かを判断する。

文字を認識できる場合は、ステップS 1 4 0へ移行してこの画像データを画像メモリに保存し、OCR処理部50による上述のOCR処理を行う。文字を認識できない場合は、ステップS 1 3 0へ移行する。

ステップS 1 3 0において、ズーム率が最大か否かを判断し、ズーム率が最大であると判断した場合は、これ以上正確な文字情報を抽出することができないと判断する。そこで、ステップS 1 3 2へ移行し、文字及び画像領域を統合した情報として出力する。

ステップS 1 3 0において、ズーム率が最大でないと判断した場合は、追加ズームとして、所定率だけズームをアップさせる。この場合に、ステップS 1 3 6として追加ズームのフラグをオンする。さらに、ステップS 1 3 8へ移行し、文字認識が可能になったか否かを判断する。文字認識が可能な場合は、ステップS 1 4 0へ移行し、この画像データを保存してOCR処理を行う。ステップS 1 3 8において、文字認識ができないと判断した場合は、ズーム率が最大か否かを判断する処理へ戻り、ステップS 1 3 2、あるいはステップS 1 3 4の処理へ移行する。

ステップS 1 4 0におけるOCR処理が終了すると、追加ズームフラグのオン又はオフを判断するステップS 1 4 2へ移行する。追加ズームフラグがオフしている場合は図3AにおけるステップS 1 1 2へ戻り、上述した処理を繰り返す。追加ズームフラグがオンしている場合は、ステップS 1 4 4以降におけるカメラ140のヘッドの首振り操作を行う。

ステップS 1 4 4として、カメラヘッドの横方向への首振りを開始する。これにより、ステップS 1 4 6として、画像が横方向に移動する。ステップS 1 4 8として、画像の移動量を判定する。画像の移動量を示すベクトル量が、画像フレームを移動させている横方向の90%未満である場合は、ステップS 1 4 4へ戻ってカメラヘッドの首振り操作を繰り返す。ベクトル量が画像フレームを移動させている横方向の90%以上である場合は、ステップS 1 5 0へ移行する。

そして、横方向の端部を検出したか否かを判断し、検出した場合はステップS 1 5 2として横方向の端部検出フラグをオンする。

横方向の一方の端部を検出しない場合は、ステップS 1 5 4へ移行して文字画像の有無を判断する。文字画像が存在すると判断した場合は、ステップS 1 4 0へ戻り、このときの画像データを画像メモリに保存し、OCR処理を行う。

文字画像が存在しないと判断した場合は、ステップS 1 5 6へ移行し、カメラヘッドを横方向における他方の端部まで移動する。

ステップS 1 5 8として、カメラヘッドの縦方向への首振りを開始する。ステップS 1 6 0として、画像が縦方向に移動する。ステップS 1 6 2として、

画像の移動量を判定する。画像の移動量を示すベクトル量が、画像フレームを移動させている縦方向の90%未満である場合は、ステップS158へ戻ってカメラヘッドの縦方向の首振り操作を繰り返す。ベクトル量が画像フレームを移動させている縦方向の90%以上である場合は、ステップS164へ移行する。

そして、縦方向の端部を検出したか否かを判断し、検出した場合はステップS166として縦方向の端部検出フラグをオンする。

縦方向の一方の端部を検出しない場合は、ステップS168へ移行して文字画像の有無を判断する。文字画像が存在すると判断した場合は、ステップS140へ戻り、この画像データを画像メモリに保存し、OCR処理を行う。

文字画像が存在しないと判断した場合は、ステップS170へ移行し、図5のステップS350における原稿設置を操作者に促すシーケンスを実行する。

従来のような、スキャナを用いて原稿上を走査して読み取り、OCRによって文字情報を得る方式では、スキャナが走査するのに時間がかかり無駄が生じる。

上述した本実施例によれば、スキャニング機能を有しない簡易な撮影ユニットや一般的のビデオカメラ等の映像撮影装置で撮影した動画映像データを入力ソースとして、静止画像を識別し、得られた静止画像から文字情報を抽出して文書データとして保存し、加工することができる。

よって、走査時間を要し効率の低いスキャナを用いることなく、動画撮影により高速に原稿を読み取り、かつ読み取った動画映像信号に含まれる文字情報を取り出して文書データ化し、再利用や鮮明な印刷が可能となる。

製本された書籍から文書データを生成する場合は、本発明の原稿読み取り装置に書籍を置いて、操作者あるいは公知の自動頁捲り装置を用いて頁を捲っていく、捲る途中で所定時間静止するようにして撮影を行う。これによれば、厚手の原稿であっても頁捲りを行っていくだけで撮影が可能であり、通常の複写機を用いる場合のように、原稿の見開き毎に原稿台に裏向きに押当てて、複写の度に開始ボタンを押すことなく高速で連続的に静止画を取り込むことができる。よって、これまで多大な時間を要していた書籍の電子データ化を促進し、省スペース、省容量化を実現することができる。

また、原稿が文字情報のみから成るものでなくとも、画像領域を除いた文字領域のみを識別し、OCR処理を行って文字変換して文字情報を得ることができる。

従って本発明によれば、スキャナ等の既存の静止画生成用装置を不要とし、汎用のビデオカメラのように、制約の小さい非定型な動画生成装置を用いて文書情報を抽出することが可能である。

ここで、画像の統合は、例えば特開平11-134352号公報、特開平11-69288号公報等に開示されている、静止画像をパノラマ画像に合成する手法を用いて行ってもよい。

上述したズーム機能を用いる場合には、1つの画面を複数のブロックに分割することになるので、次に取り込むべき画面が存在することになる。分割して画面を取り込む際に、一つの画面に存在する文章を連続した状態で取り込む必要がある場合、例えば以下の2つの手法が考えられる。

(1) 取り込んだ画像にOCR処理を施す前に、重ね合わせ部（例えば、右の端部、下の端部）を基準に重ね合わせて1枚の画像とする。

(2) OCR処理をそれぞれの画面に施した後、重ね合わせ部（右の端部、下の端部）における文書データを照合し、この重ね合わせ部で重複した文書データを消去しながら行を結合する。

また、抽出した静止画像に含まれる文字領域にOCR処理を施して文字情報を生成する場合、この文字情報の形式は、例えばフォント形式、文字サイズに関する情報が付属した、例えばテキストコードとしてもよい。

広域識別の際に記憶した各領域の位置情報に基づいて、文字情報と非文字領域（グラフィック領域）とを合成して画面を統合する場合、画像データの形式は、例えばjpeg形式、tiff形式等としてもよい。

次に、ユーザとサービス提供者とを、例えばインターネット等のネットワークにより接続した場合の構成について説明する。

静止画像に含まれる文字領域から文字情報を抽出する際に、OCR処理が必要となる。ところが、この処理には通常高精度なOCR処理演算部が必要であり、簡易な携帯型装置として実現することは困難である。

そこで図8に示されたように、ユーザが有するパーソナルコンピュータ80a、80b、…と、サービス提供者90が有するセンタサーバとが、電話回線94や携帯通信端末96等によって、インターネット92を介して接続されたネットワーク構成を構築する。

ユーザは、携帯型ビデオカメラ等を用いて撮影した動画映像を、各々のパーソナルコンピュータ80a、80b、…を用い、インターネット92を介してサービス提供者のセンタサーバ90に送信する。サービス提供者は、送られてきた画像情報に対してセンタサーバ90を用いてOCR処理を行い、文字情報を生成してインターネット92を介して各パーソナルコンピュータ80a、80b、…に返送する。このようなサービスシステムとして構築することもできる。

このようなシステムとすることで、各ユーザは高価なOCR処理装置を導入することなく、動画に含まれた文字情報を得ることができるので、コストの低

減が可能である。

(a) サービスを受けるときの操作手順

ユーザがビデオカメラにより撮影して作成した動画映像信号に対し、パソコン用いて動画ファイルを生成し、インターネットにダウンロードする場合について、図9を用いて説明する。

先ず、ステップS220として、ユーザがインターネットにアクセスし、ステップS222として、動画映像信号から文書データ作成のサービスを提供しているサービスセンタのサイトに入る。

ステップS224として、ユーザがこのサービスを受けるのが初回である場合、ユーザのサービス使用登録を行う。

ステップS226として、ユーザ名、ID番号、ユーザパスワードを入力し、サービス提供者のセンタサーバで確認する。

確認が終了すると、以下のステップS228における処理へ移行する。

先ず、希望サービス（文書データの作成）の選択を行う。

変換リソースとしての映像再生時間を入力する。

操作内容として、例えば印刷のみ、文字情報への変換のみ、印刷及び変換の両方等のいずれかを選択する。

印刷を希望する場合は、その印刷物の送付方法として、郵送、FAX等のなかからいざれかを選択する。

郵送を希望する場合、郵送する前に予めFAXにて送付を通知することを希望するか否かを選択する。

文書データの形式として、例えばテキストデータ、アドビ社PDFファイル、各種ワープロソフトファイルのいずれかを選択する。また、文書データを保存する記憶媒体の種類を選択する。

送付先の住所として、登録済みの住所、あるいは別の住所の指定を行う。

次に、ステップS230として、希望サービスの課金明細と合計算定の表示を行う。

ステップS232として、内容変更の有無の確認を行う。

変更が有る場合は、ステップS234として上記希望サービスの選択（ステップ230）に戻る。変更が無い場合は、次のステップS236へ移行する。

ステップS236として、サービス提供者側において、センタサーバのデータ格納場所を開放する。

ステップS238として、ユーザが動画ファイルをインターネット上にアップロードする。

ステップS240として、サービス提供者側がデータを受け取ったことを確認する。

ステップS 2 4 2として、サービス提供者がユーザに対し、データを受領した旨を表示する。

ステップS 2 4 4として、サービス提供者がこの受け取った動画ファイルを文書データに変換して出力する。

ステップS 2 4 6として、ユーザ側の希望サービスに応じて、FAX送信の希望があった場合は、サービス提供者がFAX送信を行う。

ステップS 2 4 8として、ユーザ側から、出力内容の可否をFAXでサービス提供者に送信し内容の確認を行う。

ステップS 2 5 0として、ユーザ側の希望サービスに応じて、印刷物と保存メディア等の郵送を行う。

(b) ユーザのサービス使用登録に関する処理

ステップS 2 6 0として、ユーザがインターネットを介して、サービス提供者のサービス(ローカル)サイトにアクセスする。

ステップS 2 6 2として、初回の場合、ユーザ登録を行う。

ステップS 2 6 4として、ユーザの身元を確認することが可能なクレジットカードの番号等の問い合わせを行う。これは、ユーザがサービスに対する代金の支払いを履行しない場合に請求先を確保するために行う。

ステップS 2 6 6として、登録費用と登録維持費とを必要とする場合、これらの費用の支払形態の決定を行う。

ステップS 2 6 8として、ユーザに関する情報を記録し、パスワードをユーザに送信する。

ここで、ユーザに対する課金方法であるが、例えばユーザの登録時に指定したクレジット口座等の決済機関;又はインターネット接続業者を介しての請求、あるいは直接ユーザに請求書を送付等が考えられる。

上述のようなインターネットを用いたサービスの提供によれば、次のような効果が得られる。OCR処理装置は、高精度なOCR処理演算部が必要であり、安価で携帯に好適な装置として構成することが困難である。このような高価な装置を、たまにしか利用しないユーザが購入するのは負担が大きすぎる。

そこで、ユーザはOCR処理装置を備えずに、OCR処理装置を有するサービス提供者に画像情報から文字情報への変換処理を依頼する。即ち、ユーザが動画映像を撮影し、転送可能な画像データを生成してインターネットを介してサービス提供者に送信する。サービス提供者は、転送されてきた画像情報をOCR処理を施し、抽出した文字情報を電子データ化したファイルとしてユーザに返送するサービスを提供する。これにより、OCR処理装置という高価なハードウェアを複数のユーザ間で共有することとなり、装置の稼働率が向上すると共に、ユーザが負担するコストを低減することができる。

上記実施例はいずれも一例であって、本発明を限定するものではなく、本発明の技術的範囲内において、様々に変形することが可能である。

例えば、上記実施例ではユーザとサービス提供者とをインターネットで接続しているが、インターネットに限らず何等かの通信ネットワークで接続されていればよい。

また、動画映像から文字情報を抽出し、文書データをインターネットを介してユーザに送付するサービスでは、ユーザが指定した印刷処理可能なステーションに文書データを直接送信して印刷物を出力してもよい。